

Cloud Computing Service for Managing Large Medical Image Data-Sets Using Balanced Collaborative Agents

Raúl Alonso-Calvo, Jose Crespo, Victor Maojo, Alberto Muñoz,
Miguel García-Remesal, and David Pérez-Rey

Abstract. Managing large medical image collections is an increasingly demanding important issue in many hospitals and other medical settings. A huge amount of this information is daily generated, which requires robust and agile systems. In this paper we present a distributed multi-agent system capable of managing very large medical image datasets. In this approach, agents extract low-level information from images and store them in a data structure implemented in a relational database. The data structure can also store semantic information related to images and particular regions. A distinctive aspect of our work is that a single image can be divided so that the resultant sub-images can be stored and managed separately by different agents to improve performance in data accessing and processing. The system also offers the possibility of applying some region-based operations and filters on images, facilitating image classification. These operations can be performed directly on data structures in the database.

1 Introduction

Different medical imaging devices and techniques used in clinical routine in many settings, such as CT, MRI, PET, X-Rays, etc., generate a huge quantity of information every day. Moreover, continuous advances in the optical resolution of medical capturing instruments and sensors imply that image sizes are growing, introducing larger storage requirements.

Raúl Alonso-Calvo · Jose Crespo · Victor Maojo · Miguel García-Remesal · David Pérez-Rey
DLSIIS & DIA - Biomedical Informatics Group,
Facultad de Informática, Universidad Politécnica de Madrid, Campus de Montegancedo,
Avda Montepríncipe, S/N, 28660 Boadilla del Monte, Spain
e-mail: ralonso@infomed.dia.fi.upm.es

Alberto Muñoz
School of Medicine, Universidad Complutense de Madrid, Spain

Building effective storage and indexation systems for medical images poses a challenging problem. The standard approach Picture Archiving and Communication Systems (PACS) aims to address and solve important issues like robustness, scalability, interoperability, and simple access to data. For this purpose there are several projects that use image databases with Content-Based Image Retrieval (CBIR) techniques.

Different kinds of CBIR have been proposed. CBIR systems usually utilize low-level features (such as, for example, color histogram techniques and texture analysis) for performing image matching. In this group we can include well-known systems like QBIC [7], Virage [3], VisualSeek [11] and Blobworld [4]. However, medical images and their contents are usually linked to semantic information such as patient data, date, image type, and additional information (such as the body parts shown in the image). By using such textual information, software systems can refine content-based image searches.

In this context, storing and managing this medical imaging information becomes a crucial task for most medical organizations. Besides, integrating this multimedia information with other textual sources is useful for different aspects related to clinical workflows, such as diagnosis, therapeutic decisions, research and teaching.

In this paper we present a distributed multi-agent prototype system capable of storing and indexing very large image collections. This approach extends previous work on our group in ontology-based database integration of structured [2] and unstructured sources [10], [8] and in image processing [5], [6].

Distributing large image databases is commonly used for improving scalability and systems response. A distinctive aspect of our work with respect to other systems previously mentioned is that, in those other systems, data distribution is limited to spreading images to different servers, whereas in our system an image can be divided so that image parts can be stored in two or more different agents. This aims to improve performance.

This paper is organized as follows. In the next section we describe the methods used to create our multiagent system for storing large collections of images. Next, we present and discuss the results obtained for querying and retrieving information from stored images in the system. Finally, a conclusion section ends the paper.

2 Methods

2.1 Cloud Computing Service

As is well known, cloud computing [9] is intended to offer an interface to use virtual resources accessible on demand, and they can be controlled by the service owner. A goal of our work is to offer a cloud computing service for storing, analyzing and retrieving medical multimedia information. Institutions and medical professionals that do not have enough storage and computing resources could manage medical data through applications built on top of these types of services. This kind of Virtual PACS based on the Cloud is different to other previously reported systems.

The Cloud computing service is accessed by users having an authorized account. Images loaded by users are normally private, although they can be set as public if desired. By using Cloud services for accessing images, private PACS belonging to concrete medical organizations can be extended in a relatively easy manner.

2.2 Multi-Agent System Architecture

A system prototype for storing and managing images has been implemented following the agent-oriented paradigm. There are three agent roles in the system:

- (i) The Cloud Computing Service Access Point Agent is a unique agent in the system. This agent is an entry point for final registered users. It offers methods for (a) storing images in the system, (b) applying operations on stored images, and (c) querying for images in the system.
- (ii) The Resource Index Agent can be seen as a directory service. This agent contains information about location, current load, estimation function for performance prediction, and image processing status, of each working agent of the system.
- (iii) Working Agents are at the core of the system. They contain all the functionalities implemented in the system, namely extracting image low-level and semantic features, storing images and its associated information, retrieving them, and applying filters and operations.

Each working agent manages a relational database implementing the region-based datastructure created for storing images. Besides, they have methods implemented to be capable of:

- Dividing an input image. As result of this division a set of processed regions and two untreated subimages are obtained. Detailed explanation of the algorithm used can be found in our previous work [1].
- Extracting regions, low-level descriptors, relationships between regions, and textual semantic information (contained, if available, in DICOM headers) from the input image.
- Storing processed information in the graph-based data-structure. By inserting the data obtained in a relational database with the graph-based datastructure logical schema.
- Sending untreated parts of the input image, obtained in the division step, to other agents to be processed.
- Obtaining and updating information from the resource index agent for load balancing.
- Communicating to other agents to synchronize information of divided images.
- Applying filters and operations that have been implemented in the system. We use mainly region-based (or graph-based) morphological operations.

2.3 Managing Images in the Prototype: Agent Behaviors

We can distinguish three different behaviors of agents in the system depending on the kind of task that has to be performed: (i) processing and storing a new image, (ii) performing a filter or operation on an image, and (iii) querying the system for image contents (this query can be a query by image-content, or simply accessing to parts of a given image).

- (i) When a new image is introduced in the system to be stored, the input image is sent to a working agent that follows an algorithm with six main steps, depicted in Figure 1. Note: An input image is divided if its area is larger than a given threshold [1].
- (ii) When a petition of applying a filter or operation on an image is received in the cloud computing service access point agent, it spread the message to those working agents that contains sub-images of the processed image. Then those working agents initialize the requested operation in their stored sub-images belonging to the selected image. These filters and operations are directly applied to the data-structure stored in the agent local database. Distributed adaptations of some most representative morphological region-based operations has been developed the system, namely erosion, dilation, and watershed.
- (iii) When the system is queried by a final user for obtaining a part of the information stored (through the cloud computing service access point), the query is propagated to all agents of the system (as in project Ontofusion [2], [10]). Each agent returns information matching with the user query restrictions that are stored in its database. Note that results are not necessarily entire images. All results are returned by the cloud computing service access point agent to the final user.

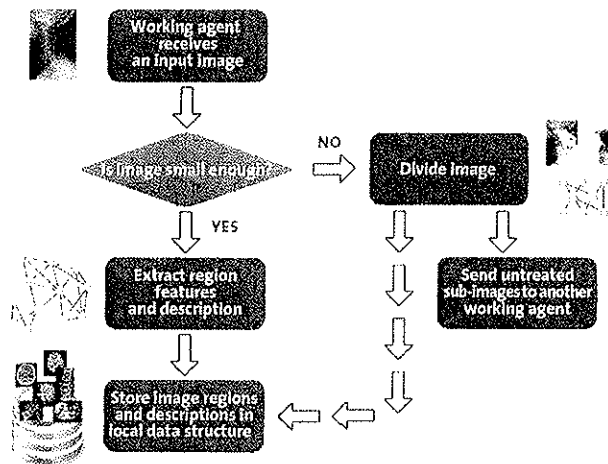


Fig. 1 Storing an image in the prototype

3 Evaluation and Discussion

We have conducted various image retrieval experiments using a set of images stored in the system. As illustrative cases that correspond to three selected images of different sizes, we have displayed in some figures execution times for three images: Image 1 is 1000x1000, Image 2 is 3000x3000, and Image 3 is 5000x5000. Results in seconds for some example queries in the system are shown when the number of agents (which are deployed in different servers) is increased from 1 to 3. Images are divided among the available agents. When more than one agent is involved, the execution time is the maximum of all agents.

Figure 2 shows the time reduction in a crop operation (where a 300x600 pixel rectangle is retrieved from the image, with the corresponding label of the region to which it belongs).

Finally, Figure 3 visualizes the execution time of a more complex image multi-criteria query. Particularly, the query has retrieved all regions having (a) an area less than a constant, (b) an intensity value greater than a threshold, and (c) both a major and a minor axis smaller than 90% the image size (such a multi-criteria query is often used to eliminate background regions).

As can be observed the time needed for executing these queries decrease when the number of agents in the system increases. The data division and distributed management among different agents, has decreased the time needed by database searches and operations.

Many of the existing database CBIR use distributed systems for storing images. This image distribution in different servers reduces the time needed for database queries. Besides this kind of parallelization, our system can also divide input images and store subimages in a distributed manner to further reduce the time for queries. This paper has focused on this second, novel type of reduction. Another important feature of our system is that operations are applied directly to databases.

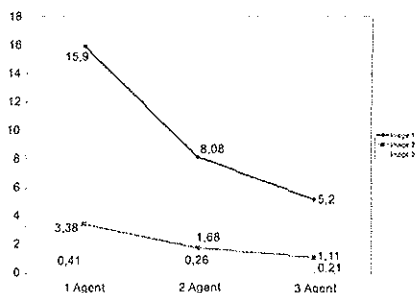


Fig. 2 Crop operation

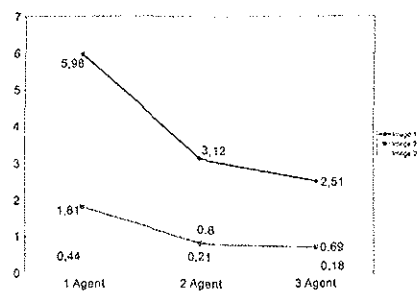


Fig. 3 Multi-criteria query

4 Conclusions and Future Work

In this paper we have presented a scalable multi-agent system prototype capable of storing very large image collections. The system offers methods for managing,

analyzing and querying images by content through a cloud computing service. The system uses a distributed region-oriented data-structure that can be stored, in a distributed manner, in one or more databases. Some experimental results have been provided that show the performance increase in normally used operations when the number of agents is increased.

First, we plan to extend the image operators implemented by adding new ones, such as region merging. Another planned improvement is to create methods in the cloud computing service for offering intelligent streaming of image contents, which can be used to visualize complex medical image data.

Acknowledgements

This work has been supported in part by "Ministerio de Ciencia e Innovación" of Spain (Ref.: TIN2007-61768).

References

1. Alonso-Calvo, R., Crespo, J., García-Remesal, M., Anguita, A.: On distributing load in cloud computing: A real application for very-large image datasets. In: International Conference on Computational Science, vol. 1(1), pp. 2663–2671 (2010), doi:10.1016/j.procs.2010.04.300
2. Alonso-Calvo, R., Maojo, V., Billhardt, H., Martín-Sánchez, F., García-Remesal, M., Pérez-Rey, D.: An agent- and ontology-based system for integrating public gene, protein, and disease databases. *Journal of Biomedical Informatics* 40(1), 17–29 (2007)
3. Bach, J., et al.: Virage Image search engine: An open framework for image management. In: *SPIE Storage and Retrieval for Image and Video Databases*, vol. 2670, pp. 76–87 (1996)
4. Carson, C., Thomas, M., Belongie, S., Hellerstein, J.M., Malik, J.: Blobworld: A system for region-based image indexing and retrieval. In: Huijsmans, D.P., Smeulders, A.W.M. (eds.) *VISUAL 1999*. LNCS, vol. 1614, pp. 509–517. Springer, Heidelberg (1999)
5. Crespo, J., Serra, J., Schafer, R.W.: Theoretical aspects of morphological filters by reconstruction. *Signal Process.* 47(2), 201–225 (1995)
6. Crespo, J., Schafer, R.W., Serra, J., Gratin, C., Meyer, F.: The flat zone approach: a general low-level region merging segmentation method. *Signal Process.* 62(1), 37–60 (1997)
7. Flickner, M., et al.: Query by Image and Video Content: The QBIC System. *IEEE Computer* 28(9), 23–32 (1995)
8. Maojo, V., García-Remesal, M., Billhardt, H., Alonso-Calvo, R., Pérez-Rey, D., Martín-Sánchez, F.: Designing new methodologies for integrating biomedical information in clinical trials. *Methods of Information in Medicine* 45(2), 180–185 (2006)
9. NIST definition of Cloud Computing, vol. 15 NIST (2010), <http://csrc.nist.gov/groups/SNS/cloud-computing/> (cited April 2011)
10. Pérez-Rey, D., Maojo, V., García-Remesal, M., Alonso-Calvo, R., Billhardt, H., Martín-Sánchez, F., Sousa, A.: Ontology-based integration of genomic and clinical databases. *Comput. Biol. Med.* 36(7-8), 712–730 (2006)
11. Smith, J., Chang, S.: Blobworld: VisualSeek: A fully Automated content-based image query system. In: *ACM International Conference on Multimedia*, pp. 87–98 (1996)